

Figure 1

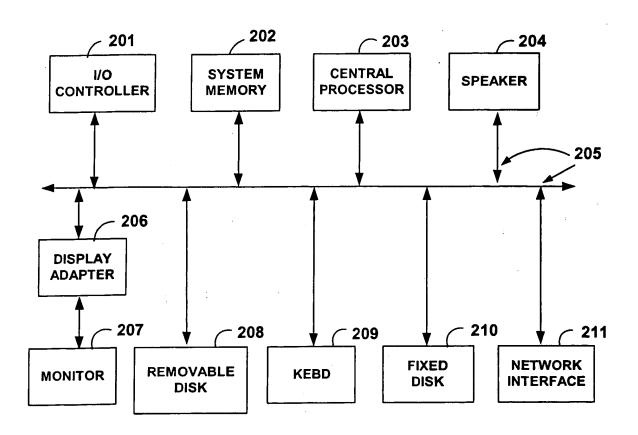


Figure 2

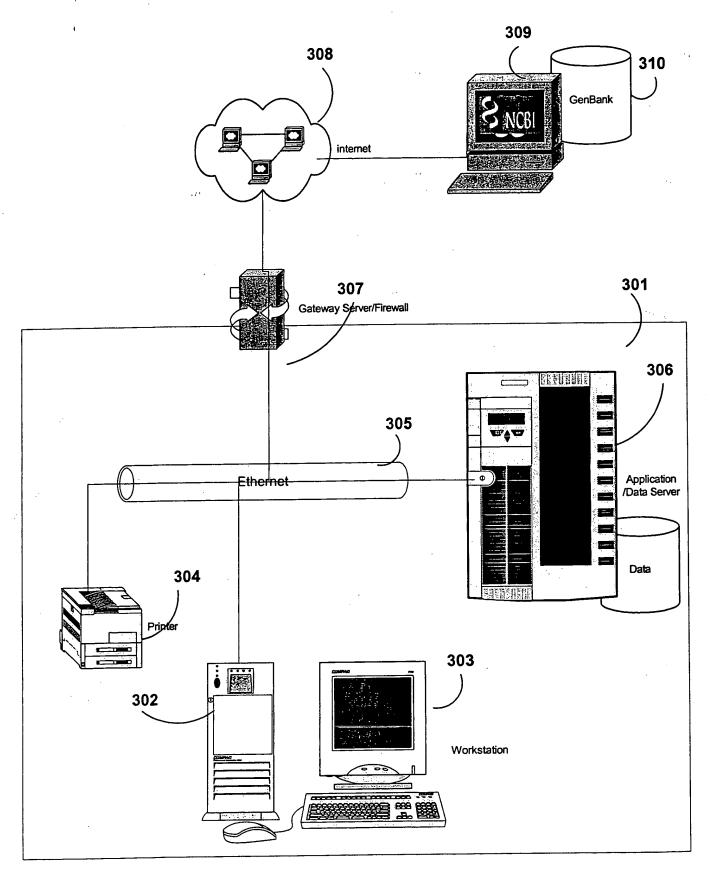


FIGURE 3

BEST AVAILABLE COPY

Sequence Collection

• Sequences and annotations collect from a variety of sources.

Sequence Analysis

- Aligned to the draft assembly of the genome.
- Low quality bases trimmed from ESTs.
- Polyadenylation sites detected.

Seed Cluster Creation

- · Seed clusters created for each UniGene Cluster.
- Additional seed clusters are created with potential full-length mRNAs not in UniGene.

Genomic Based Subclustering

Seed clusters are subclustered based on the contig.

Sequence Based Subclustering

• Sequences are subclustered in transcriptome space.

Orientation Based Subclustering

• EST read directions, CDS annotations, consensus splice sites, and polyadenylation sites are used to predict the orientation of the subcluster. The subcluster is resubclustered if substantial conflicts exist.

Picking Probe Selection Regions

- For each subcluster, regions for probe selection are selected from either the consensus or exemplar sequences.
- Multiple regions may be choosen due to alternative polyadenylation sites.

Prioritizing Sequence Selection Regions

 Regions for probe selection are prioritized based on the quality of sequence and annotation. A1 A2 A3

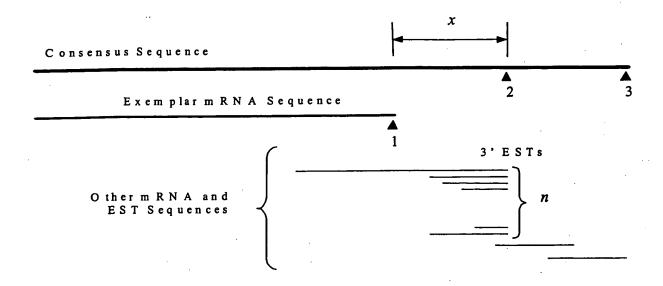


FIGURE 6

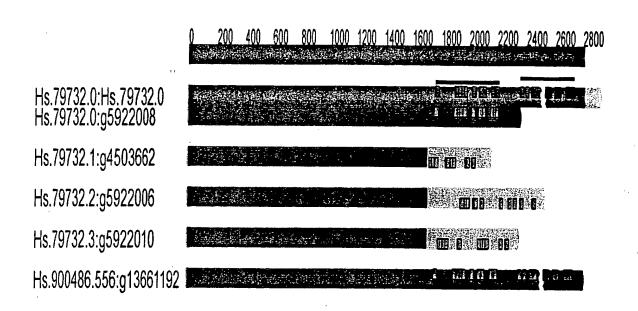


Figure 7